

Prakyat Prakash

+1 585-202-8784 | prakyat02@gmail.com | [linkedin.com/in/prakyatp](https://www.linkedin.com/in/prakyatp) | github.com/Prakyatp | Rochester, NY

EDUCATION

Rochester Institute of Technology

Rochester, NY

Masters in Computer Science

Aug 2024 – Dec 2026

- **Relevant Coursework:** Data Structures and Algorithms, Computer Architecture, Machine Learning, Foundations of Artificial Intelligence, Big Data, Advanced Computer Vision

PUBLICATIONS

Pizzi, J.R., Adhikari, I., **Prakash, P.**, Miyamoto, H., & Cui, F. (2026). **Machine learning-based determination of sex-related bladder cancer biomarkers.** *Frontiers in Bioinformatics*, 6. <https://doi.org/10.3389/fbinf.2026.1794098>.

Carson, C.W., Umesh Nagalakshmi, S., Adhikari, I., Freewoman, J.M., Pizzi, J.R., **Prakash, P.**, & Cui, F. (2026).

Intrinsic DNA codes govern distinct modes of nucleosome-transcription factor interactions. *Nature Communications*. <https://doi.org/10.1038/s41467-026-74448-4>

EXPERIENCE

Graduate Research Assistant

May 2025 – Present

Machine Learning Research Assistant

Rochester, NY

- Engineered an end-to-end Python-based ETL pipeline to extract, clean, transform, and validate 24.4M raw peptide-spectrum matches from large biological databases, producing 361,789 experimentally verified phosphosites using 35 parallel workers.
- Developed scalable data transformation workflows and structured data models to integrate heterogeneous biological datasets, implementing data quality checks, deduplication, and FASTA verification to ensure consistency and reliability.
- Designed data engineering solutions for large-scale biological data processing, optimizing parallel execution, database integration, and reproducible data pipelines for downstream machine learning applications.

Acinonyx Technologies private limited

Jan 2025 – Apr 2025

AI Engineer - Contract

Remote, NY

- Developed Python-based data processing and transformation pipelines for LLM-powered RAG applications, improving code generation accuracy by 25% through retrieval optimization and prompt engineering.
- Implemented structured data retrieval, database integration, and automated evaluation workflows using LangChain, reducing model output errors by 30%.
- Applied software development life cycle (SDLC) principles, Agile methodologies, Git-based version control, and CI/CD practices to deliver scalable AI data solutions.

PROJECTS

PTM Dataset Extraction Pipeline | *NumPy, UniProt, PyTorch* | *Live Dashboard*

Dec 2025 – March 2026

- Reconstructed phosphorylation and lysine acetylation training datasets from 5 mass spectrometry repositories sourced from the DeepMVP/PTMAtlas compendium, processing 24.4M raw peptide-spectrum matches into 361,789 tryptic-verified unique phosphosites across S, T, and Y residues.
- Replaced the field-standard assumption that all non-positive sites are negative by extracting true negatives from experimentally observed unmodified residues.
- Corrected trypsin-cleavage bias in acetylation data via two-stage K-position downsampling, reducing terminal-K over-representation from 33.2% to 3.6%.

Cross-Border Healthcare Navigator | *Python, RAG, Neo4j, LLMs, Vector DB*

Ongoing

- Built Python-based ETL and data transformation workflows to collect, clean, and structure healthcare cost information across 10+ countries and 30+ medical procedures for analytical and retrieval applications.
- Designed a Neo4j NoSQL graph database and structured data models to represent healthcare providers, treatment costs, recovery periods, and patient decision pathways across 5+ touchpoints.
- Developed cloud-based data solutions using AWS Lambda, API Gateway, and database-backed RAG services, reducing estimated healthcare research time by 40%.

TECHNICAL SKILLS

Languages: Python, SQL (PostgreSQL), Java, C++, JavaScript

Data Engineering: ETL Pipelines, Data Transformation, Data Integration, Data Warehousing, Data Lakes, Apache Spark

Cloud and Databases: AWS (S3, DynamoDB, Lambda, API Gateway), Neo4j, PostgreSQL

Software Engineering: SDLC, Agile Development, Docker, Git, CI/CD, REST APIs

Machine Learning: PyTorch, TensorFlow, Scikit-learn, Pandas, NumPy, LangChain, RAG